DTO

VACE
Video Analysis
Content Extraction

Text Recognition Evaluation

Rangachar Kasturi, Dmitry Goldgof
Rajeev Sharma
Rachel Bowers, John Garofolo

Evaluation Team
Padmanabhan Soundararajan
Vasant Manohar
Yan Qiu
Matthew Boonstra
Valentina Korzhova

Annotation Team
Harish Raju
Hankyu Moon
Shubha Prasad
Carmen Mazzant

RT 2006

USF
UNIVERSITY OF
SOUTH FLORIDA

NIST
National Institute of
Standards and Technology

VideoMining

Overview

- Introduction
- Task Definition
- Annotations
- Metrics
- Scores
- Conclusions

Introduction

- DTO funded
- Coordinated by NIST
- Univ. of South Florida (USF) and VideoMining (VM) team
- ViPER: Annotation tool developed by Univ. of Maryland (UMD)
- Data distribution managed by LDC/NIST

How we have Evolved

- VACE-I
  - Detection
  - Tracking (given first frame reference)
- VACE-II
  - Detection and Tracking on a bigger dataset (50/50)
  - No reference given
  - Pilot evaluation on Text Recognition
    - BBN/SRI (single participant)

Text: Task Definitions

- Detection Task: Spatially locate the blocks of text in each video frame in a video sequence
  - Text blocks (objects) contain all words in a particular line of text where the font and size are the same
- Tracking Task: Spatially/temporally locate and track the text objects in a video sequence
- Recognition Task: Transcribe the words in each frame, including their spatial location (detection implied)
- Currently use Broadcast News data

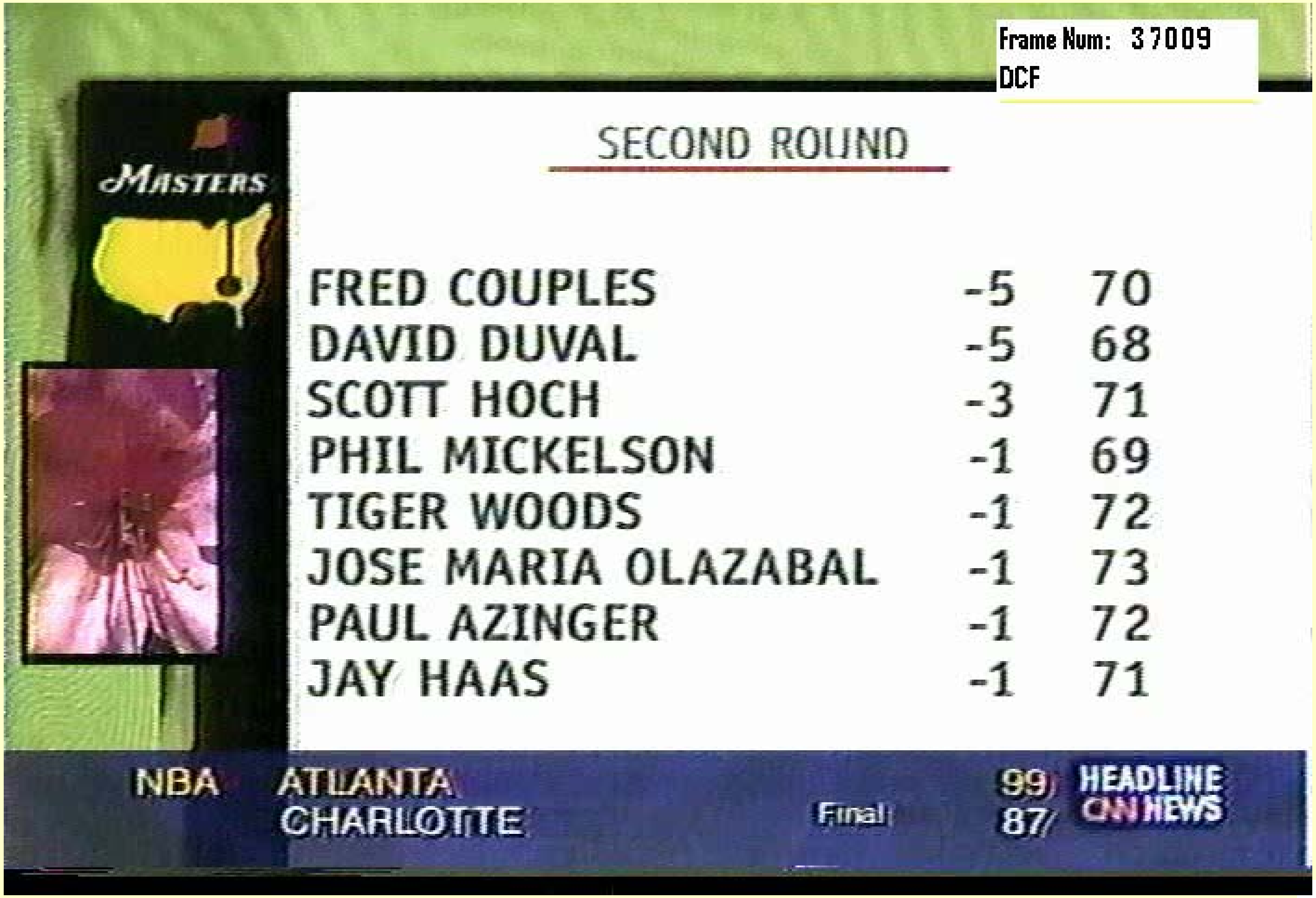
Task Definition Highlights

- Annotate oriented bounding rectangle around text objects
- Detection and Tracking task
  - Line level annotation with IDs maintained
  - Rules based on similarity of font, proximity and readability levels
- Recognition task
  - Word Level (IDs maintained)
- Documents
  - Annotation guidelines
  - Evaluation protocol
- Tools
  - ViPER (Annotation)
  - USF-DATE (Scoring)

What types of Text are evaluated?

- Evaluate only the most easily readable text (to establish a baseline at a high level of inter-annotator agreement)
  - Type = graphic (no scene text)
  - Readability = 2
  - Logo = false
  - Occlusion = false
  - Ambiguous = false
    - Exclude scrolling (ticker), dynamic text (scoreboard)
  - Case insensitive and punctuation ignored

Sample Annotation Clip



Frame Num: 37009

DCF

SECOND ROUND

FRED COUPLES	-5	70
DAVID DUVAL	-5	68
SCOTT HOCH	-3	71
PHIL MICKELSON	-1	69
TIGER WOODS	-1	72
JOSE MARIA OLAZABAL	-1	73
PAUL AZINGER	-1	72
JAY HAAS	-1	71

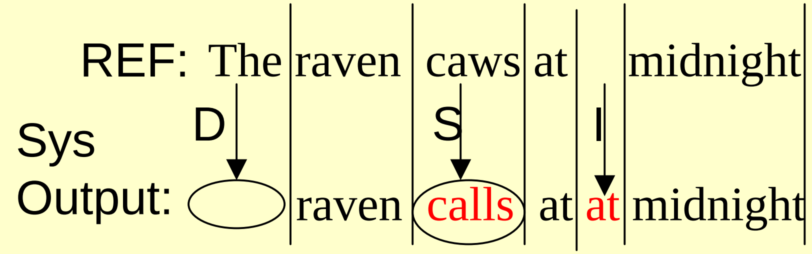
NBA ATLANTA 99

CHARLOTTE Final 87

Metrics

- Spatially map system output detected words to reference words, then compare the strings for mapped words
  - An unmapped word in system output incurs an Insertion (I) error
  - An unmapped word in reference incurs a Deletion (D) error
  - A mapped word with a character mismatch incurs a Substitution (S) error

$$\mathrm{WER} = \frac{(\mathrm{I}+\mathrm{D}+\mathrm{S})}{(\text{Total \# Words in Ref})}$$



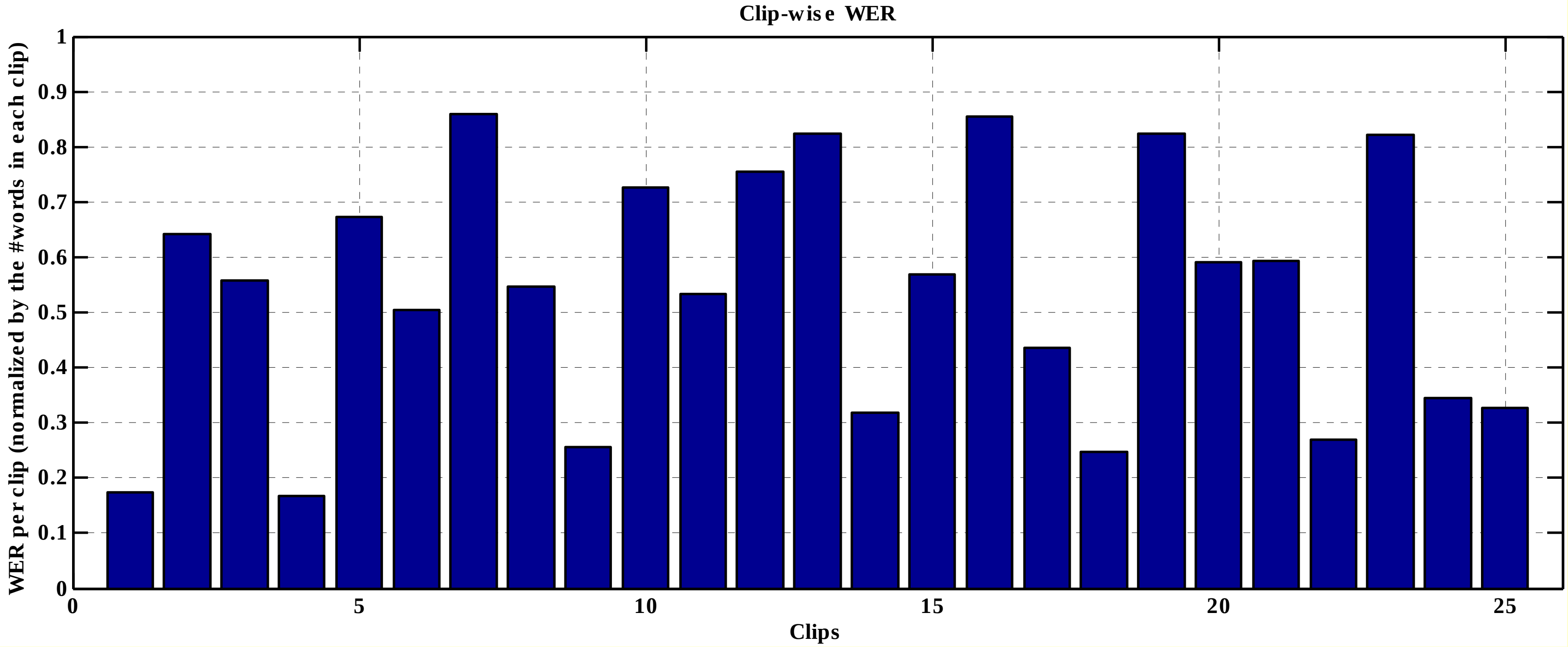
WER = (1 + 1 + 1)/5 = 3/5 (60%)

- Errors are accumulated over entire test set
- Also generate: Character Error Rate

Datasets

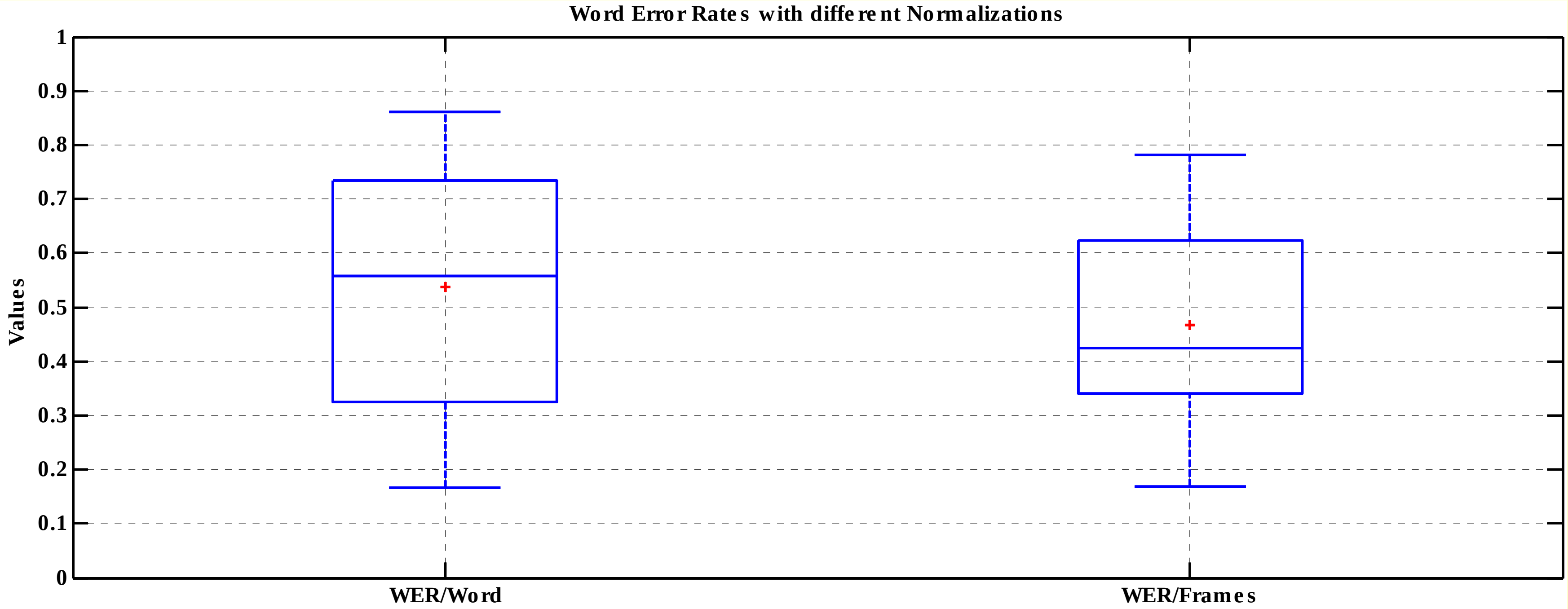
- Broadcast News (1998 TDT corpus)
- Training/Dry Run Development Set
  - 5 Clips
    - 14.5 minutes
    - 1181 words
- Evaluation Set
  - 25 Clips
    - 62.5 minutes
    - 4178 word objects
    - 68,738 word frame instances

Individual Clip WER



Scores (Word Error Rate)

Participants: SRI/BBN



WER	CER
0.4233	0.2823

Discussions

- Harder to recognize if the word occurs rarely?
  - Further analysis needed to verify this
- Total number of words in test set: 68,738
- Total number of System generated words: 54,628
  - Detection and Recognition errors

Conclusions

- Successful Pilot project
- Ported ASR metrics to Video evaluations
- Good Baseline result

Future Work

- More challenging forms of text
  - Non—Broadcast News domains (MRoom, Surveillance) and foreign BNews (Chinese + Arabic)
  - Handwritten, dynamic, scrolling, harder to read, scene?
- Text object structural and semantic grouping
  - Will enable non—bag–of –words NL processing

Acknowledgements

- SRI/BBN
- NIST: John Garofolo, Rachel Bowers
- UMD ViPER team

Thanks
&
Questions